DOCUMENT RESUME

ED 234 077 TM 830 561

AUTHOR Higgins, James E.; And Others
TITLE Measurement and Evaluation in Indiana Public Schools.
INSTITUTION Indiana State Univ., Terre Haute. Curriculum Research and Development Center.; Indiana State Univ., Terre Haute. Evaluation Assistance Center.
PUB DATE Feb 83
NOTE 40p.; Paper copy not available due to small print.
AVAILABLE FROM Curriculum Research and Development Center, Statesman Towers West, School of Education, Indiana State University, Terre Haute, Indiana 47809.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Administrators; *Computer Assisted Testing; *Educational Testing; Elementary Secondary Education; Evaluation Methods; Measurement Techniques; *Public Schools; Self Evaluation (Individuals); *Standardized Tests; *Teacher Made Tests; Teachers; *Test Wiseness
IDENTIFIERS *Indiana

ABSTRACT

This report discussed the following questions in regard to testing, measurement and evaluation in the Indiana public schools: (1) Where are we? (2) Is "where we are" different from where we perceive we should be? (3) Do teachers feel competent in testing? (4) How are teachers' skills viewed by their "superiors," the central office administrators? and (5) Where would educational assistance staff time and energy be best served? It is felt that the results of this data collection can be used to develop additional inservice thrusts, and also to identify preservice needs that can and should be met by faculty members in the various teacher-training colleges and universities within the state. Opinionnaires were mailed to approximately 1200 teachers, 900 principals, and 300 superintendents--with a return rate of 348, 418, and 164 respectively. Specific analyses of selected data are addressed in the following chapters of this monograph: (1) classroom testing, (2) standardized testing, (3) pupil test-taking skills and the testing environment, (4) the computer in testing, and (5) an administrator's self-appraisal. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED234077

MEASUREMENT AND EVALUATION IN INDIANA PUBLIC SCHOOLS

by

James E. Higgins

James E. Isaacs

Frank W. Jerse

Jerry A. Summers

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

[X] This document has been reproduced as received from the person or organization originating it.

[ ] Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Nelson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EVALUATION ASSISTANCE CENTER

TM 830 561

CURRICULUM RESEARCH AND DEVELOPMENT CENTER

SCHOOL OF EDUCATION, INDIANA STATE UNIVERSITY • TERRE HAUTE

THE EVALUATION ASSISTANCE CENTER

The Evaluation Assistance Center at Indiana State University was established through a grant provided by the Lilly Endowment Foundation. The work of the EAC centers around any and all aspects of measurement and evaluation of pupil achievement.

The EAC provides services on a contractual basis to all interested school corporations, in-state and outside the boundaries of Indiana. These services include conducting needs assessments, assisting in the formulation of measuring instruments to assess pupil learning, and working with public school personnel on improving their respective skills in the measurement and evaluation of their pupils' achievement.

The EAC is housed in the Curriculum Research and Development Center.

THE CURRICULUM RESEARCH AND DEVELOPMENT CENTER

The Curriculum Research and Development Center of Indiana State University provides school systems the opportunity to secure aid, encouragement, and cooperation in curriculum development projects. It coordinates the participation of University personnel engaged in curriculum work, provides information concerning curriculum development, and initiates and sponsors curriculum research projects. It is the contact point where public schools initiate inquiries regarding curriculum and acts as a vehicle for communication between elementary and secondary schools and the University. Although the CRDC operates as an agency of the School of Education, it represents all departments of the University engaged in curriculum development projects.

John C. Hill
Interim Dean, School of Education

James E. Higgins
Director

This monograph is sponsored jointly by the
Evaluation Assistance Center
and the
Curriculum Research and Development Center

MEASUREMENT AND EVALUATION IN INDIANA PUBLIC SCHOOLS

CURRICULUM RESEARCH
AND DEVELOPMENT CENTER

SCHOOL OF EDUCATION, INDIANA STATE UNIVERSITY ● TERRE HAUTE

The material for this bulletin was written by:

James E. Higgins
Professor of Education
Indiana State University

James E. Isaacs
Doctoral Fellow
Indiana State University

Frank W. Jerse
Professor of Education
Indiana State University

Jerry A. Summers
Professor of Education
Indiana State University

School of Education
Publications:

Dorothea Nelson
Managing Editor

Carole Gustafson
Copy Editor

Sue Lightfoot
Technical Assistant

Published by:

Evaluation Assistance Center
and
Curriculum Research and
Development Center
School of Education
Indiana State University
Terre Haute, Indiana 47809

March 1983

TABLE OF CONTENTS

INTRODUCTION
by James E. Higgins 1

CLASSROOM TESTING
by James E. Higgins 7

STANDARDIZED TESTING
by Frank W. Jerse 11

PUPIL TEST-TAKING SKILLS AND THE TESTING ENVIRONMENT
by Jerry A. Summers 15

THE COMPUTER IN TESTING
by James E. Higgins 19

ADMINISTRATOR'S SELF-APPRAISAL
by James E. Isaacs 21

CONCLUDING REMARKS
by James E. Higgins, Frank W. Jerse, and Jerry A. Summers 25

APPENDIX 29

INTRODUCTION

James E. Higgins

The Evaluation Assistance Center was founded in 1979 through a grant from the Lilly Endowment, Incorporated. The Center is staffed by faculty and graduate students in the School of Education at Indiana State University, Terre Haute. The major activities of the Center consist of offering consultant services to public school personnel to assist them in developing and furthering their skills in measuring and evaluating pupil achievement. Another contribution made by the Center is the development of monographs concerning specific aspects of measuring and evaluating. Previous monographs are *Indiana History Test: Fourth Grade*, *Leadership for Effective Evaluation of Learning*, and *Using Measurement and Evaluation to Promote Learning*.

Other Center activities include sponsoring of measurement and evaluation conferences; making presentations at local, state, and national meetings; and conducting of research. Whereas members of the staff have conducted numerous "action" research projects with public school personnel within Indiana, the research reported in the following chapters of this monograph represents an endeavor of much greater significance with implications more far-reaching.

Problem

Throughout the existence of the Evaluation Assistance Center, staff members have been called upon to provide inservice on testing, both standardized and classroom. Public school personnel from around the state, and in some instances from neighboring states, have made telephone inquiries to Center staff concerning standardized test selection and utilization. There is an increased awareness of, and interest in, assisting pupils in developing test-taking skills. The microcomputer is being introduced to schools and there is some interest in its being used in a measurement and evaluation capacity, although the interest is not matched with developed expertise at the time of this writing.

Because of these forces, the staff of the EAC determined it necessary to get a picture of "where we are" in testing, measurement, and evaluation. Also, is "where we are" different from where we perceive we should be? There were numerous other questions to be answered: Do teachers feel competent in testing? How are teachers' skills viewed by their "superiors," the central office administrators? If the EAC had to focus on just a few areas, rather than using the "shotgun" approach that has been used in the previous three years, responding to <u>all</u> needs, where would staff time and energy be best served?

It is felt the results of this data collection can be used to develop additional inservice thrusts, and also to identify preservice needs that can and should be met by faculty members in the various teacher training colleges and universities within the state.

Procedures

Beginning in January 1982, the staff of the Evaluation Assistance Center spent much time in the development of three different opinionnaires--one each for teachers, principals, and central office administrators. The instruments (see Appendix) were designed to detect congruencies and discrepancies in educators' perceptions. For instance, teachers were asked to respond to items on the basis of "where they are" and "where they should be." Therefore, it was possible not only to determine the degree of skill teachers perceived themselves to have, but also to determine whether they felt the need for even more skill. Principals were asked to indicate the degree of skill they perceive teachers should have and then to evaluate themselves based on their feelings of what their own skills should be. In general terms, the data from principals yields a picture of what principals perceive teacher roles should be and what they perceive principal roles should be in the measurement and evaluation process. Central office administrators were asked to indicate the degree of skill and/or familiarity they perceived teachers and principals to have and compare that to where they feel teachers and principals should be. An additional section was provided on the central office opinionnaire to allow them to evaluate their own strengths and needs.

When looking at the instruments, it is obvious that each is quite lengthy; measurement and evaluation is far more complex than most people realize. Recognizing that the longer an opinionnaire the less chance for getting responses, it was decided to do all that was possible to promote a high percentage of return. This included having the opinionnaires professionally printed in color, using the Center logo. A personalized, addressed and signed cover letter accompanied each opinionnaire. Return postage was provided.

Opinionnaires were mailed to approximately 1200 teachers, 900 principals, and 300 superintendents. The sample of teachers and principals was randomly selected by computer, this service being provided by the Department of Public Instruction. All superintendents received opinionnaires, and in most instances additional opinionnaires were included for the superintendents to distribute to their curriculum directors and assistant superintendents.

The return, in the eyes of EAC staff members, was good, especially considering the length of the instruments. Thirty-two percent of the teachers (N=348) returned completed opinionnaires. The percentage of responses from principals (N=418) and central office administrators (N=164) far exceeded that. Therefore, the sample size upon which conclusions are drawn is statistically adequate.

The Sample

More specifically, the sample can be broken down into various subcategories. For instance, teachers were asked to indicate what grade or grades they presently teach. It was found that:

1. A total of 49 teachers indicated they teach in grade three or below, including kindergarten.

2. There were 42 intermediate grade (4-6) teachers responding.

3. Eighty-three teachers indicated they teach in junior high or middle schools.

4. Teachers indicating they teach in grades seven through twelve were considered to be jr./sr. high school teachers. There were 41 of them.

5. One hundred twenty-nine teachers responded that they teach high school (9-12 or 10-12).

6. Four teachers failed to provide this information on their opinionnaires.

Just what, specifically, did these teachers teach? Most of the elementary teachers indicated they taught all subjects in self-contained classrooms. The junior high and senior high teachers taught math (46 in all), science (36), language arts (49), social sciences (29), the humanities (33), physical education and health (20), with the rest being "scattered" through special education, business, industrial arts, and home economics.

Principals were asked to indicate the number of years they have spent in the principalship. It was felt that the number of years in the position might be related in some way to the manner in which they responded to the opinionnaire items. Of the 418 principals responding, 56 indicated they had three or fewer years experience. One hundred thirty had been a principal for 4-9 years, while the majority (223) had ten or more years experience. Nine principals failed to respond to that particular opinionnaire item.

In looking at the enrollments in the schools of the responding principals, approximately 75 percent of them were in schools with enrollments of 200-800 pupils. Only 22 principals indicated they were from schools with less than 200 enrollment, while 76 principals indicated the enrollments of their schools were quite large (better than 800).

The central staff responding to the opinionnaire occupied at that time the following positions: There were 66 superintendents, 43 assistant superintendents, and 33 directors of curriculum. There were 22 additional responses where the information was not given or the position held was not identifiable with any of the above three groups. Better than 75 percent of these respondents were from school corporations having less than 500 certified staff. A few (18) had between 500 and 1000 certified staff in their corporations, while eight respondents indicated more than 1000 such employees.

The staff of the Evaluation Assistance Center was curious as to how many respondents had taken a graduate course in educational measurement and evaluation. Better than 80 percent had, with the majority of those who had not coming from the ranks of teachers. Thus, when the data are analyzed and conclusions are drawn concerning needs existing in testing, measurement, and evaluation, it will likely be pointed out that these needs exist for reasons other than lack of opportunity to gain information concerning the topic.

Data Analysis

Specific analyses of selected data are addressed in the following chapters of this monograph. However, it is necessary here for examples to be provided to allow the reader a better grasp of how the data were anlayzed. For instance, the following two statements served as items on the opinionnaire:

Familiarity you have					Familiarity you should have			
4	3	2	1	When mastery testing is appropriate	4	3	2	1
4	3	2	1	When it is best to "grade on the curve"	4	3	2	1

Teachers were asked to mark the left column on the basis of the degree of familiarity they feel they have with whatever the item on the opinionnaire refers to, and the right column was to be used to indicate the degree of familiarity they should have. The numbers referred to the following:

4--thorough
3--adequate
2--minimal
1--none

The teachers' mean average response for the item concerning mastery testing was 2.68 for the left column and 3.29 for the right column. The following conclusions are drawn:

1. Teachers feel less than adequately prepared (a mean response of 3.0 would indicate teachers, overall, feel adequate in this area.)

2. Teachers feel their degree of preparation (familiarity) should be more than adequate (3.29 exceeds the "adequate" rating of 3.0).

3. The difference between the two responses is significant (t-tests were run on each pair of responses on the opinionnaire). When t values exceed 2.58, the significance is beyond the .01 level ($p < .01$). Conclusions and/or recommendations concerning inservice and preservice work for teachers could also be given, and those will be dealt with in later chapters.

Central office administrators were asked to respond on the basis of the degree of familiarity they feel teachers and principals have and the degree of familiarity they feel teachers and principals should have. This may cause the reader some confusion, as it has the EAC staff, since principals and teachers are "lumped together." In looking at the data, the writers of this monograph have done their best to correctly interpret what has been found and to make appropriate inferences and recommendations. For instance, the mean responses for the item concerning mastery testing were 2.37 and 3.34, respectively. This would suggest that central office personnel perceive the teachers (and/or principals) being less prepared than the teachers perceive themselves being. Are there further conclusions to be drawn? Yes, and they are discussed later in this monograph.

Principals were asked to respond to the item on the basis of the degree of familiarity teachers should have (left column) and the degree of familiarity they feel they, the principals, should have (right column). The mean responses for the item concerning grading on the curve yielded the following results:

1. Teachers should be more than adequately prepared in this aspect of measuring and evaluating (3.49 mean response).

2. Principals should also be more than adequately prepared (3.29 mean response).

This suggests that principals, although they are not doing the actual grading of pupils/students in schools, still perceive a great need to have familiarity with various grading practices.

The following chapters have been written by individual members of the Evaluation Assistance Center staff and reviewed by all staff members prior to publication of this monograph. Each is an attempt to share the extensive data from this study in such a way as to allow the reader to grasp the present status of teachers and administrators in measuring and evaluating pupil acheivement. If, after reading this monograph, additional questions concerning the data arise, write or call the Center for additional information and/or clarification.

CLASSROOM TESTING

James E. Higgins

The opinionnaire items concerning classroom testing concerned four major teacher tasks:

1. making the test blueprint,
2. constructing the test items,
3. analyzing the test results, and
4. sharing those test results with parents of the students.

Teachers were asked to indicate their degree of familiarity with each of these tasks and to compare that to the degree of familiarity they feel they should have. For the purposes of this chapter, "familiarity" is considered to be synonymous with "skill," thus a particular degree of familiarity indicated by teachers would be interpreted as the degree of skill those teachers feel they have or should have. Principals were asked to respond on the basis of the degree of skill teachers should have and also the degree of skill they, principals, should have. Central office administrators were asked to give their opinion of where teachers presently are (skill level) and where they feel teachers should be.

Generally speaking, teachers perceive their own skills in classroom testing to be adequate or nearly so, although they do suggest a need for significantly higher degrees of skills. However, central staff perceive teachers to be not nearly as well prepared as the teachers themselves indicate they are.

It is not possible within the confines of this monograph to share all the data gathered, nor would it be worthwhile to do so. Rather, the contents of this chapter will share findings and implications that are significant, statistically and/or educationally.

The Test Blueprint

Teachers indicate a degree of unfamiliarity in making the test blueprint. Possibly this is due to a lack of understanding of what the blueprint actually is. More commonly known as a "table of specifications," its use is much the same as what a blueprint is to a contractor when he is building a house. It is reasonable to suggest that teachers have not had a great deal of experience in actually planning a test thoroughly. Far too many tests are developed shortly before they are to be given. Surprisingly, some of those tests are fairly high quality, but a blueprint increases the chances. Making a test without a blueprint is similar to the practice of writing a term paper prior to writing the outline.

Central office administrators' mean response (2.27) to this item on the opinionnaire suggests they view teachers and principals to be far less than adequately prepared (barely "minimal") in this area. Regardless of whose rating is more accurate, the fact remains that teachers' skills should be upgraded here, likely through inservice. Such a development of skills will result in tests that have better balance (this increases

content validity) and higher reliability and, therefore, greater accuracy of measurement of pupil achievement.

Test Types

Central staff perceive teacher and principal skills in actual test construction to be less than adequate. The highest mean response of the central administrators when asked to assess where teachers presently are was 2.76 (multiple-choice and short-answer testing). As might be expected, teachers view themselves to be more skilled than that. Teachers, on the whole, view themselves as being adequately skilled in giving tests that are multiple-choice, matching, short-answer, and completion. True-false test construction is barely less than the adequate level (2.98). Elementary teachers generally view themselves as being less skilled than do secondary teachers, but that difference is not significant.

There were two types of test items that teachers indicated themselves to be less than adequately prepared to use: classification and essay. It is interesting to note that central administrators also viewed teachers to be weakest in classification and essay testing. As might be expected, the ratings given teachers by central staff were lower than the teachers had given themselves, but nevertheless central staff appears to have a fairly good "read" on what the greatest needs of teachers are in classroom testing.

Since multiple-choice testing may be used to measure different "levels" of learning, that is, memory, comprehension, inference, it is good to note that teachers feel adequately prepared to use such tests. However, the three other test types (matching, short-answer, and completion) are used to measure knowledge at the recall or memory level only. If teachers rely heavily on these three types of test items, does this also suggest that their teaching is directed toward the recall of specific facts and information? If teachers teach in such a way as to expect students to learn beyond the memorization of isolated facts, is it not also reasonable to expect them to test the depth of student learning? The answers to these questions have important implications for quality (high or low) schooling.

As mentioned earlier, teachers rate themselves less than adequate to use classification and essay tests. Since classification tests are not dealt with in great detail in many measurement and evaluation courses or textbooks, it is fair to suggest that most respondents were somewhat unsure as to what was being asked on the opinionnaire. However, such could not be the case when considering the essay. And, not only did teachers indicate a lack of preparation to use the essay, but they also indicated a relatively low degree of need to have skills in using the essay test. Just what might this imply?

Much is being written in the literature concerning high school graduates' inabilities to express themselves in written form. Do teachers feel the situation is beyond repair, so to speak, and therefore elect to use tests other than essay to measure pupil learning? This would be a detriment to the development of student skills in writing. Or, do teachers view the practice of essay testing as being too difficult to objectively evaluate students? Or, is it likely to be neither of the above? Could it be that class sizes have increased significantly during the past few years due to the economic forces acting upon the schools? If this is so, teachers with large classes would rightfully elect to use tests other than essay for the reason that essay testing with large groups is inefficient. For whatever reasons, teachers consider themselves to be inadequate in their preparation to use the essay test and, when compared to other test item types, they are not too concerned about this inadequacy.

Test Analysis

The opinionnaire item on analyzing test results from classroom tests was not specific at all. It could have been interpreted to mean comparing test content to the instructional objectives, using statistics to analyze each item, comparing pretest and posttest results as would be common practice in a criterion-referenced setting, or merely having another teacher look over the test items prior to giving the test. For whatever reason, teachers view their greatest need in teacher-made testing to be this area--test analysis. Whereas their mean rating of their present abilities (2.93) would suggest they feel themselves to be practically adequate, they rate where they "should be" at 3.52. Principals and central staff administrators also indicate teachers should have a high degree of skill in this area. However, central administrators view teachers' present skills to be quite low (2.34).

If the central administrators are correct in their assessments of where teachers actually are at this time, this would represent an enormous "spread" between what is and what should be, thereby suggesting there are immediate needs for inservice in the area of analyzing classroom test results. It is important to interpret this data with caution, however, because teachers do not view themselves to be nearly as inadequate as their administrative counterparts suggest they are. Because of this, there is potential for conflict if inservice is automatically instituted on the basis of the perceptions that administrators have of teacher skills.

Sharing Results with Parents

Overall, teachers perceive themselves to be adequately prepared to share test results with parents. However, when breaking down the data to compare elementary teachers' responses to those of jr./sr. high school teachers, the difference in perceived skills is significant. Elementary teachers view themselves to be more than adequately prepared (3.36) while teachers of older students view themselves to be less than adequately prepared (2.86). It is interesting to note that when analyzing the responses of these two groups of teachers on all opinionnaire items dealing with classroom testing, this is the only one where there is a significant difference in their mean responses.

Why is this so? One reason could likely be that elementary teachers conduct parent conferences far more often than do secondary teachers, thereby they have more experience and have continual "on-the-job training." Therefore, it would be easy to brush this difference aside and say, "Oh, well, that's just the way it is." However, when looking at teacher responses concerning the degree of familiarity they should have in sharing results with parents, the secondary teachers give a rating of 3.37. This suggests they feel there is a great need to possess skills in meeting with parents. In looking at the responses of principals to this item, when they were asked to indicate the degree of familiarity teachers and principals should have in sharing test results with parents, elementary and secondary principals responded with mean ratings of 3.61 and 3.64, respectively. This certainly indicates they see the sharing of results with parents to be highly desirable, and there is no significant difference between these elementary and secondary perceptions.

The implications from this data appear to be obvious. Communication with parents should increase, especially at the junior high and secondary school level. An open house once or twice a year is not enough. Report

cards that go home every six weeks or nine weeks are not enough. Parents, on the whole, would likely welcome this increased communication between school and home. With teachers and principals holding similar views on sharing test results with parents, there would appear to be nothing stopping this from occurring--the sooner the better.

STANDARDIZED TESTING

Frank W. Jerse

Each year school systems throughout the nation invest a considerable amount of time and money in standardized testing programs. The purpose is to obtain information for decision making by a variety of persons. These decision makers may be students and parents who are making course selections, career decisions, and college choices. They may be teachers who are seeking to understand their students better or to plan developmental or remedial activities for individuals or groups. They may be principals who need to evaluate programs and provide leadership for both short- and long-term developments. Or, they may be counselors or other pupil personnel workers who use the results to help their clients understand themselves better. Central office staff members may be using the results to evaluate overall curriculum and teaching effectiveness, to identify areas of strength and weakness, and to initiate program development. Superintendents may be using the results as they endorse recommendations for staffing and programs or as they develop reports to school boards and answers to critics.

The above reflect only a partial listing of the purposes of standardized testing programs. The degrees of proficiency and efficiency with which standardized testing results are used to meet purposes such as these vary considerably among school systems. These differences may be due to varying goals and needs of populations served or they may be due to the competencies and attitudes of the school personnel who are in positions to use the data. It was to this last consideration that this study was addressed.

Nine items on this opinionnaire were developed to sample the thinking of teachers, principals, and central staff personnel concerning perceived and desired familiarity with selected aspects of standardized testing. These aspects included development of the program (two items) and utilization of the findings for classroom or curricular applications (two items). Response from teachers, principals, and central staff personnel were then examined and compared to determine their perceptions of what the situation was and what it should be.

Standardized Testing Program Development

How familiar are teachers and principals with the justification of the standardized testing program? How familiar should they be? As a group, teachers see themselves as having somewhat above minimal but below adequate familiarity. The mean response was 2.54. However, this item showed a significant difference of expression between elementary and secondary teachers; elementary teachers showed a mean of 2.71, where as secondary teachers showed a mean of 2.47. Similarly, when asked the level of familiarity teachers should have with the justification of the testing program, elementary teachers indicated they should be well above the adequate level (3.28), but secondary teachers indicated slightly below adequate (2.96) as being an acceptable level of familiarity. These significant differences between elementary and secondary teachers appeared over and over again throughout the items dealing with standardized testing. Elementary teachers tend to see themselves as having and needing greater familiarity than do secondary teachers. Principals tend to see teachers as needing slightly less than adequate (2.89) familiarity with program

justification but they demanded a much higher level of familiarity for themselves (3.66). Apparently they perceive themselves as needing to provide leadership in program development. Central office staff perceive teachers and principals as having below adequate proficiency (2.54), but as needing a significantly higher level (3.24).

A second item asked for self-ratings concerning selecting a test. Teachers show a significant difference between their perceived familiarity level (2.39) and the level where they felt they should be (3.36). The responses of the principals are similar to those of the previous item, showing they expect to exercise leadership. Once again they feel that they should have a degree of familiarity approaching thoroughness (3.66) while asking teachers for an adequate degree (3.07). Central office staff responses rated both groups as having well below adequate familiarity (2.34) and as needing an above adequate degree (3.38).

Classroom and Curricular Application

How familiar are teachers with the use of test results in individual classrooms? Elementary school teachers report themselves as having almost an adequate degree of proficiency (2.91) and as needing significantly more (3.52). Their principals agree with the last rating; their rating of what teachers should have is identical to that of the teachers' rating for themselves (3.52). Elementary principals' rating for themselves (3.24) shows they expect more of teachers than of themselves in this area.

The picture for secondary teachers and principals is quite different. Secondary teachers see themselves as having significantly less familiarity than do elementary teachers (2.55) and as needing significantly less (3.19). Secondary principals sharply disagree. Their rating of the degree of familiarity secondary teachers should have is 3.60 (approaching thoroughness) as compared to the teachers' rating of 3.19 (slightly above adequate). Central office ratings agree with those of elementary teachers and both groups of principals as to the familiarity level which should be attained (3.59).

The item which dealt with the use of test results for schoolwide curriculum changes showed different responses than the previous item. Principals at both levels think they need significantly more familiarity with using test results for curriculum change than they expect of teachers (3.79 vs. 3.22). The principals' ratings approach the thorough category. Apparently they see themselves as providing leadership. The expectations of central office staff also are for a high degree of familiarity (3.47). Interestingly enough, significant differences appear between responses of elementary and secondary teachers, both for perceived level of familiarity (2.93 vs. 2.58) and needed level of familiarity (3.52 vs. 3.20). Once again, elementary teachers show greater involvement with the testing program.

Sharing Test Results

The third category of items dealt with the sharing of results with students, parents, and the public. Results differ to a considerable degree, depending upon which of the three groups is being considered.

With respect to sharing results with students, both elementary and secondary teachers respond as having well above minimal, but below adequate, familiarity (2.70) and below what they should have (3.23). Both levels of principals agree with the teachers as to the desired levels, as did central

office personnel. However, data for the item dealing with sharing test results with parents shows a number of significant differences. Elementary teachers report their familiarity level to be slightly above adequate (3.10); this is significantly higher than the reports by secondary teachers as to their familiarity level (2.44). Elementary teachers also show significantly higher mean scores concerning the level of familiarity they should have (3.50) than do secondary teachers (3.17). This may reflect the demands made on them in their work. Elementary principals agree with elementary teachers as to the desired level of familiarity teachers should have. However, secondary principals have significantly higher expectations for secondary teachers (3.46) than the teachers do for themselves (3.17). Both groups of principals have expectations approaching thoroughness for their familiarity with sharing test results with parents (3.62); they expect somewhat less from their teachers (3.40).

The need of familiarity for sharing with the public provides an interesting view of role expectations for teachers with respect to this topic. Teachers and principals, alike, rate both existing teacher familiarity and expected teacher familiarity as being below the adequate level, with one exception. Only elementary school teachers think that teachers should have adequate familiarity (3.01). Elementary and secondary teachers show ratings barely above the minimal level (2.31 and 2.01) for their perceptions of where they are. Secondary teachers feel that they need below adequate familiarity (2.72), and elementary and secondary principals agree (2.64 and 2.69). Principals, however, rate this as quite necessary for them, showing an overall mean rating of 3.51. Also, of interest are the ratings by central office staff, showing 2.28 for their estimate of where teachers and principals are, and 3.09 of where they should be. (Unfortunately, the opinionnaire asked central office staff to rate teachers and principals as a unit and it is not possible to determine precisely whom they were rating.) Principals, on the other hand, rate themselves as needing a high degree (3.51) of familiarity with sharing of results with the public. Apparently, this task,which has public relations aspects, is one which is deemed to be a task for administrators rather than teachers.

Technical Proficiency

The final category of items studied the degree of familiarity respondents had with the use of technical information in using test results. One item dealt with the interpretation of test scores; the other was concerned with the confidence which could be placed in test results. Once again, elementary and secondary teachers show significant disagreements. Elementary teachers report adequate familiarity with interpretation of test scores (3.04) as opposed to the below-adequate rating reported by secondary teachers (2.62). The need for familiarity with this skill also shows significant differences between the two groups of teachers. Elementary teachers indicate they should have familiarity approaching thoroughness (3.58), while secondary teachers say that their level should be slightly above adequate (3.18). The same differences are evident for the item dealing with confidence in test results. Elementary teacher ratings of familiarity they think they have (2.79) and what they think they should have (3.45) differ significantly from the ratings by the secondary teachers (2.53 and 3.20, respectively).

Both elementary and secondary principals agree with teachers that teachers should have familiarity with test score interpretation above the adequate level (3.19 and 3.30). However, the elementary principals' ratings for teachers (3.19) is significantly below the level that teachers desired for themselves (3.58). Both groups of principals see themselves

as needing familiarity approaching thoroughness (3.63 for elementary and 3.60 for secondary principals), a level which is significantly higher than their expectations for their teachers. Central office staff members rate teachers and principals as midway between minimal and adequate for the degree of familiarity they have with both concepts (2.52 and 2.44 for the combined groups for the two items--interpretation of test scores and the confidence which could be placed in these scores), and indicate that teachers and principals should have well above average familiarity (3.38 and 3.46 for the combined groups for the two items).

Discussion

A number of general conclusions can be drawn from the data. These conclusions have implications for teachers, administrators, and other school personnel as they seek to improve the services they offer. They raise questions concerning roles and functions of teachers and principals, and they can be used as guides for inservice and continuing education programs.

Both elementary and secondary teachers report that they have somewhat above minimal but below adequate degrees of familiarity with standardized tests and the use of the results. Elementary teachers, however, rate themselves as both having and needing significantly more familiarity than secondary teachers report for themselves. These differences may be due to the amount and type of contacts they have with students, to differing definitions of the role of the teacher, to differences in demands made upon them by students and parents, or to differences in the amount of preparation for and involvement with the testing program.

Elementary and secondary school principals indicate that both they and their teachers need to increase the degree of familiarity so that it will be above the adequate level; they all agree that familiarity above the minimal level was not enough. Central office staff show general agreement with both principals' and teachers' evaluations of themselves.

Principals' ratings for themselves, on all items except two dealing with classroom use of results, show significantly higher expectations for themselves than for teachers. Apparently they see themselves as having the responsibility of leadership and demand high levels of familiarity for themselves.

Although elementary principals and elementary teachers agree on their ratings on all items except one dealing with teachers' sharing test results with the public, the secondary principals and secondary teachers differ on several items. The secondary principals have significantly higher expectations for the teachers' utilization of test results in the classroom than the teachers do. They also expect teachers to have a greater degree of familiarity with sharing results with parents. These findings for secondary school personnel raise questions as to whether there are differences in role expectations by the two groups. It may reflect a need for inservice education programs which will help teachers to function more effectively.

Finally, teachers see themselves as having low familiarity with procedures involved in sharing test results with the public. Whereas principals might agree that teachers have less than adequate familiarity, they indicate that teachers do not need an adequate degree. Elementary teachers disagree with the latter finding, indicating they feel they should have an adequate degree. These findings may be indicative of overall expectations for teachers' communications with the public. It needs to be explored.

PUPIL TEST-TAKING SKILLS AND THE TESTING ENVIRONMENT

Jerry A. Summers

Four items were identified on the EAC opinionnaire on measurement and evaluation which deal specifically with the concept of pupil test-taking skill and the testing environment. The items are teaching test-taking skills, preparing students for test taking, physical conditions appropriate for testing, and psychological factors affecting test results. Efforts will be made in this chapter to (1) define the meaning of the concept of test-taking skill and the test environment, (2) present the findings of the appropriate status study items in comparative fashion, and (3) make conclusions and recommendations as appropriate for teacher education, inservice education, school curriculum, and school practices.

The Concept of Test-Taking Skill and the Test Environment

Clearly there is no way of knowing how each of the respondents personally interpreted any one item on the assessment instrument. Test-taking skills and testing environments embrace several subject-matter areas, and there are probably as many different interpretations as there are individual responses. In the classical sense, however, the concepts have been interpreted such that they are independent of cognition. Test-taking skills have traditionally focused on concepts such as the following:

1. Dealing with test anxiety
2. Learning how to improve scores through guessing including the identification of
   a. Context clues
   b. Specific determiners
   c. Pattern responses
   d. Syntax hints
   e. Testing bias
   f. Root words
   g. Systematically illogical distractors
3. Testwiseness
   a. Reading with comprehension
   b. Identification of main objectives
   c. Speculating with precision
4. Testing environment
   a. Maximizing comfort and efficiency
   b. Minimizing confusion
   c. Timing
5. Practice in taking tests
   a. Use of answer sheets
   b. Practice in testing type: multiple choice, true-false, short answer, essay, etc.

A review of the literature in testing, measurement, and evaluation strongly supports the notion that achievement on tests of all types--teacher-made and standardized--can be improved with a systematic effort to help sutdents develop test-taking skills. As suggested before, however, it is unlikely that the respondents considered most of these variables as they formulated their responses to the status study opinionnaire items.

Teaching Test-Taking Skills

When the responses of teachers, principals, and central staff regarding teaching test-taking skills are compared, a discrepancy is identified which suggests that teachers believe they should have considerably more skill in this area than they presently have. Their present skills are represented as less than adequate (2.63), and they desire to be rated as adequate or above (3.33). Central staff are found to agree with the teachers; they suggest that at present the skills of teachers in this area are even worse than the teachers believe they are (2.33).

Principals suggest that teachers should have more skill in this area than they, the principals, should have.

Preparing Students for Test Taking

In a comparison of responses regarding preparing students for test taking, it is found that teachers report that their skills in this area are adequate (3.29), but they also believe they should be even higher (3.64). Principals and central staff appear to agree substantially in the skill level they believe teachers should have. Central staff rank the skills that teachers have as less than adequate (2.74).

Physical Conditions Appropriate for Testing

In comparing the responses for teachers, principals, and central staff regarding physical conditions appropriate for testing, it is found that teachers believe their skill levels are adequate in this area (3.25); central staff disagree and rank the skill level of teachers as less than adequate (2.80). Teachers, principals, and central staff agree that teachers need to be proficient in this area.

Psychological Factors Affecting Test Results

When the responses of teachers, principals, and central staff regarding psychological factors affecting test results are compared, it is found teachers believe that their skills are adequate (3.00); central staff, again, disagree, and they mark the skill level that teachers presently have as less that adequate (2.80). Teachers, principals, and central staff agree that this is an important area and that teachers should be well prepared in dealing with psychological factors affecting test results.

Conclusions and Recommendations Regarding Test-Taking Skills

Conclusions and recommendations for this unit are divided into four areas. Attempts have been made to make recommendations for (1) teacher education, (2) inservice education, (3) school curriculum, and (4) school practices.

Teacher Education. The evidence presented in this chapter suggests clearly that the area of teaching test-taking skills is not adequately covered in teacher education institutions across the State of Indiana. Neither teachers nor central staff believe that teachers are adequately prepared in this area. Further, all three groups of respondents (teachers, principals, and central staff) indicate that the skills of teachers should be considerably higher than they presently are.

The skills of teachers are ranked as adequate by the teachers for preparing students for test taking, physical conditions of testing, and psychological factors affecting testing. Central staff disagree on the last item; they suggest that the skills of teachers are less than adequate in the psychological factors affecting test results.

The following recommendations are directed toward teacher education institutions.

1. Teacher candidates should be given a fairly rigorous accounting of test-taking skills in the preparatory program. Such an effort should make a pupil's test score more reflective of actual achievement and possibly improve comprehension in the more general concepts of the curriculum.

2. Clinical components of the teacher education program should require competencies or evidence of performance for teacher candidates in teaching test-taking skills. Such an effort would not only improve pupil achievement, it would also direct the candidate's attention to other elements of testing, measurement, and evaluation.

3. The subject matter for teaching test-taking skills should be directed to colleges, schools, or departments of education for instruction for students in their preclinical phase. Efforts should be made to assist teacher candidates in translating classroom "theory" into "clinical practice."

4. Areas of greatest need appear to be related to the test-taking skills rather than the physical or psychological factors affecting testing. This researcher's interpretation of the data suggests that teachers and administrators, by and large, accept these factors as "given" in the natural repertoire of teaching skills (in other words, a candidate should not have to be taught that a "hot" room results in an "uncomfortable" testing environment).

Inservice Education. With the absence of an adequate preservice emphasis on test-taking skills, the data suggest that inservice efforts to assist teachers in raising their skill levels should be effective. Further, the administrative staff at both the school and the central office levels would be supportive of such endeavors. Correlated concepts of measurement, evaluation, and research may also be improved if the attention of teachers is directed toward test-taking skills. The subject matter emphasis should focus on concepts of test-taking skills rather than physical and psychological factors affecting the testing environment.

Teachers are probably not routinely given responsibility in the presentation of all-school test results to the public. They make their

presentations to parents and other interested individuals, but they are seldom placed in a position to defend (or more correctly to explain) why students did well or poorly on a given test. Such responsibilities would undoubtedly enhance interest on the part of teachers to further assist their pupils in test achievement.

School Curriculum. The task of curriculum directors in the public schools should include the responsibility for the total educational program in the schools. This includes the concepts of planning and organizing the curriculum, instructional integrity (ensuring that the objectives are taught in the classrooms), and, finally, assessment (including the evaluation of the assessment instruments to ensure that the objectives advanced were in fact the objectives that were tested). Responsibility for the academic success of the pupils in a school should be shared among all of the partners in the educational process (teacher educators, teachers, principals, curriculum directors, central administrators, and members of the board of education).

School Practices. School administrators probably are the most influential agents for change in a school system. They, more than any other individual, have the capacity to direct and encourage the development of innovative educational programs.

Principals are probably aware of these needs; they consistently rate their needs almost equal to the needs of teachers in knowing how to deal with teaching test-taking skills. Clearly, unless administrators are knowledgeable of the concept of pupil test-taking skill and the benefits that may accrue as a result of a program that develops these skills, the existence of an effective curricular or inservice model that incorporates such concepts would probably be a "chance" occurrence. Therefore, the recommendation is made that administrators--central staff as well as school principals--become familiar with the research, benefits, and programs available in the area of test-taking skills. They have the responsibility to provide whatever leadership is necessary to ensure the inclusion of these concepts in the school curriculum.

THE COMPUTER IN TESTING

James E. Higgins

The computer, more specifically the microcomputer, is just beginning to find a place in the educational setting. At the time of this study (Spring 1982), it is suggested that only about one in five Indiana schools have one or more microcomputers at hand. As a result, few teachers and administrators know much about the uses of these machines, whether those uses be for testing or any other educational endeavor. In putting together the opinionnaire, the Evaluation Assistance Center staff felt they could pretty much predict the results for this particular opinionnaire section. As will be discussed, the actual results are somewhat different from what had been expected.

The opinionnaire dealt with five different uses of the computer:

1. Using the computer to test pupils
2. Storing test data on computer disk or tape
3. Using the computer for test interpretation
4. Using the computer for data analysis
5. Retrieving pupil assessment data

In general, teachers and administrators indicate very little familiarity (skill) in using the computer or microcomputer for any of the five stated purposes. This result was expected. However, one unexpected result was that teachers, as a whole, failed to see a high degree of need for using the computer for any one of these tasks. The staff members in the Evaluation Assistance Center speculate that this result is likely due to a general lack of understanding of the capabilities of the machine rather than an apathetic view toward using computers in schools.

Other chapters in this monograph point out differences held by teachers and central office staff concerning teachers' skills in measurement and evaluation. It is interesting to note that on the five items dealing with the computer in testing, teachers and central office administrators have no significant differences in their views of where teachers presently are or where they should be; of all the data derived in the study, this is the only section where central staff do not rate teachers significantly lower than the teachers rate themselves.

However, when looking at the data collected from principals we begin to find significant differences. You will recall that principals were asked to indicate the degree of familiarity teachers should have and then to indicate the amount of familiarity they (the principals) should have. It is when comparing these sets of data that some interesting conclusions may be drawn.

Principals indicate teachers do not need too much familiarity with using the computer in testing. In fact, the data show that principals see less need for teacher knowledge (2.80) in this area than do the teachers (2.90).

Principal perceptions of where teachers should be (2.80) and where principals should be (3.06) in this area are even more startling. In every instance the difference "favors" the principals over the teachers and the differences are each significant ($p < .01$). And, the principals indicate the need for more than adequate familiarity for themselves, but not for the teachers.

Although there is no way of knowing for sure what these data mean, it is possible to speculate. Could it be that principals see little possibility of computers being available to teachers on a regular basis, thus the lower need for teachers to be "in the know"? If this is so, why are the ratings for teacher need as high as they are? At the time of this writing, microcomputers are more likely to be found in the principals' offices than in teachers' classrooms. Is this because that's where they actually belong?

Looking from another perspective, do principals see the computer as a source of power? If so, would they not wish to become familiar with its uses in testing prior to teachers actually putting the technology to work? When examining the opinionnaire items, one could agree that "using the computer to test pupils" is definitely a teacher task. Therefore, isn't it fair to expect that principals should rate teachers' needs at least as high as principals' needs on that item? But that's not what resulted. Was there a desire to retain this "power," even in the classroom?

To the results of this opinionnaire section, the author offers the following suggestions and/or recommendations:

1. Public school principals should quickly become familiar with the uses of computers, specifically microcomputers, in the educational setting. This knowledge should not be limited to, but should include, the measurement and evaluation component.

2. A cadre of teachers from each attendance center should be given released time and other types of assistance for purposes of developing their knowledges and skills, awareness, and literacy with respect to computing.

3. Indiana certification laws should require evidence of computer literacy for the procurement of teaching and administrative licenses.

4. Institutions of higher education responsible for teacher training should develop curricula to meet the above-mentioned certification requirements.

5. This research should be replicated within two years to determine what, if any, changes have taken place concerning the uses of computers in testing, measurement, and evaluation.

ADMINISTRATOR'S SELF-APPRAISAL

James E. Isaacs

Central office staff members were asked to respond to the same items as the teachers and principals. They were to indicate the degree of familiarity teachers and principals have and the degree to familiarity they felt these two groups should have. In addition the central office personnel were to evaluate themselves with respect to twelve items. It is to these items we will turn our attention in this chapter.

An examination of the twelve items (see Appendix) reveals that they are divided into two sections: Using Standardized Tests and Using the Computer (Microcomputer). Central staff persons were to indicate in the left column the degree of familiarity they have and in the right column the degree of familiarity they feel they should have. As in the previous sections, mean responses in the left column were compared to mean responses in the right column, utilizing a t-test. All comparisons were found to be statistically significant ($p < .01$).

In general the results show that these staff persons see their familiarity as less than adequate in all areas, and they seem to feel especially inadequate with regard to their familiarity with the use of the computer. They also seem to feel they should have more than adequate familiarity in all of these areas.

These feelings of being less than adequate can be attributed to a number of causes. It could be a lack of preparation in the area of measurement and evaluation. The taking of such a course is often required for administrative certification; however, the time since such a course was taken could be great. Inadequacy in uses of standardized tests could also reflect upon the quality of such a course; if so, these results could be saying that teacher training institutions need to examine what they're doing in this area. Perhaps more needs to be done to help these people in the field; this help could take the form of inservice or programs at administrative conferences.

Computers and microcomputers have burst upon the scene so suddenly that most educators have not been able to keep up with all of their potential uses for schools. The low mean responses in the degree of familiarity already possessed is very easily understood. Central office personnel feel very inadequate in this realm of knowledge. They also feel a great need to be highly familiar with the possible uses of the computer.

Teacher training institutions have not, until recently, offered courses for educators in the uses of computers. Such training needs to be offered and educators at all levels should avail themselves of these learning opportunities. Inservice programs need to be developed to help in this regard; computer firms can present these programs or they can be developed with cooperation from colleges and universities. This need will become more pronounced if Congress passes the bill allowing computer firms to give computers to every school as a tax write-off. If the training isn't available at that point, schools will have the machines and not know what to do with them, and they will suffer the same fate as the language labs.

An examination of some of the individual items also reveals implications for educators. Central office staff feel they are most familiar with using standardized tests for identifying curricular weaknesses (2.89), diagnosing curricular strengths (2.88), and placement (tracking) of students (2.87). These responses, although the highest, are still indicative of feeling less than adequately familiar. They also indicate they need to be most familiar with using these tests for diagnosing curricular weaknesses and strengths (3.55 and 3.51, respectively). This would seem to say they see this as the most important use of standardized tests. Their mean response of 3.23 with regard to using testing for tracking is the lowest in needed familiarity. Although this was one of the highest in the areas of possessed familiarity, they don't see this as one of the highest priorities for utilization of test results.

We also examined differences in responses between those who serve as superintendents and those who serve in other central office positions. We found only one item on which the difference was statistically significant: familiarity already possessed regarding using tests to diagnose curricular strengths. The superintendents had a mean response of 2.68 and the other staff had 3.00. This difference can be explained by the types of positions possibly held by these others. Some of them are undoubtedly employed as curriculum directors or assistant superintendents for curriculum and instruction. It is to be expected that they would feel more adequate in this area than a general superintendent.

In this survey, we find information regarding the roles of the various levels in education. In earlier items (19-20), the central office personnel cite teachers and principals as being very inadequate in their familiarity with sharing results with the public (2.28); at the same time, they feel that these two groups should have adequate familiarity (3.09). In items 93 and 94, they were asked to evaluate their possessed and needed knowledge of sharing program effectiveness with the general public. These items in separate parts of the opinionnaire seem to be comparable. They also feel less than adequate (2.86) in sharing with the public and feel a real need to be familiar with this use of testing (3.43). If we compare these responses in the need column, we see that central staff feel that sharing information gained from the testing program is more their role or responsibility than it is the role of teachers or principals.

Closely tied to these items is the item regarding using standardized tests to justify curricular change. Superintendents and other central staff indicate a need to be adequately prepared in this area (3.41). This also seems to reveal a role to be played by the central office. On earlier items (13-14) they indicate that teachers and principals should be familiar with interpreting results for curriculum change (3.33). Comparing these mean responses indicates a perception of roles with curricular change being seen as more of a role for the central office.

Reacting to items 90 and 92, the respondents indicate a need to know more about using standardized tests to determine teacher and principal effectiveness (3.35 and 3.28, respectively). These responses could be saying that the central office simply wants to know more about this use of testing; however, some may be interested in using test results as a means of evaluating teachers and principals. Although the test results may indicate the possibility of problems, especially if the lower performance is consistant over a period of years, there are too many other

variables affecting test results to use them to appraise job performance. This would certainly cause protests by teacher advocacy organizations.

In evaluating their knowledge of uses of the computer, the administrators feel they are strongest in using it for test score interpretation and for analyzing test results, although the knowledge was barely above minimal (2.10 and 2.05, respectively). It is these same two areas in which they feel the greatest need to be familiar (3.27 and 3.25). This would seem to say they see this as their most important use for a computer or mircocomputer.

Using the computer to make curricular projections was cited as their greatest weakness (1.79), but they indicated a need to have adequate familiarity with this use (3.17). An adequate familiarity (3.18) with utilizing the computer to score tests was desired, and again the possessed familiarity (2.03) was barely minimal. The lowest mean response for needed skill (3.14) was for data storage and retrieval, indicating this as not the highest priority for computer usage.

After examining this data, we would make the following recommendations:

1. A graduate-level course in measurement should be required of all people entering the field of administration.

2. A basic course in use of the computer should be required of all prospective administrators.

3. Teacher training institutions should take the lead in providing courses and inservice programs for educators in the field.

4. Administrators should avail themselves of all opportunities to enhance their knowledge of measurement and evaluation and computer utilization.

CONCLUDING REMARKS

James E. Higgins, Frank W. Jerse, and Jerry A. Summers

The tasks of measuring and evaluating are not new in terms of education or societies as a whole. There appears to be good evidence that the Chinese were using testing as a way of determining occupational placement, and thereby affecting social class, as long ago as 2000 B.C. It might be interesting to research whether those scoring high on such tests were selected (or directed) to be teachers or if teachers were relegated to a lower social standing as they appear to be in today's society. Massachusetts, a leader in public schooling in the colonies, following their laws of 1642 and 1647, utilized essay testing on curriculum content and religion for purposes of determining teacher "certification." Currently there are signs on the horizon that suggest there will soon be an increased use of measurement and evaluation at the university level for the purpose of screening those applying for admittance to a teacher education curriculum and, in some cases, there may be exit requirements for receiving a teaching license.

Once college and university graduates in education do reach the public schools, there are other reasons for measuring and evaluating. There is need for diagnosis of pupil strengths and weaknesses, and this may be conducted through appropriate testing. Students may be placed in one group or another (tracking) through the use of test results, more than likely test results from a standardized instrument, and such a practice is possible if decision makers fully understand the meaning of the test results. During instruction, as in the middle of a unit of work, teachers often give "spot" quizzes and carry out other less formal means of appraising pupil achievement to determine to what degree pupils are learning. This is called "formative evaluation" and likely plays a key role in success or failure of teaching. And then, of course, semester-end exams and unit tests are given to assist teachers in assigning marks or grades.

One can fairly easily document the misuse of tests in schools. For instance, the Scholastic Aptitude Test has recently come under fire from critics concerned with the effects the test results have on individuals and groups. When criticism is levied, however, it is important to differentiate between the test instruments themselves and the manner in which the test results are used. If the instrument is of poor quality, then the solution is simple: The test should not be used. If, however, the instrument yields highly valid results when used properly but those results are in practice inaccurately or inappropriately interpreted, the problem is far more serious and much less easily remedied. This cautionary note applies to classroom testing as well as to standardized testing.

Educational measurement and evaluation carries with it two major purposes, and they are not mutually exclusive. Data are gathered in an attempt to gain a better understanding of the teaching-learning situation in individual classrooms, with the goal of promoting learning. Data are also gathered in an attempt to provide direction for quality decision making that affects curriculum. With these two major foci in mind, the Evaluation Assistance Center conducted a statewide assessment of teacher and administrator skills in measurement and evaluation during the Spring of 1982 in an attempt to determine the present "state of affairs."

In brief, respondents were asked to compare their present level of skill to the level at which they perceived their skills should be. When the Center staff analyzed the results, it became apparent that where practitioners feel they are in terms of measurement and evaluation is not where those same practitioners feel they should be. This may be considered a positive finding because in some instances, the teachers and administrators surveyed indicate that their present skills are at the "adequate" level, but still they see a need to further sharpen those skills.

Working alongside public schools and colleges and universities is the Indiana Department of Public Instruction. In the late 1970s in an attempt to encourage local school systems' educational effectiveness, IDPI mandated that all school corporations within Indiana should develop and/or revise local curricular objectives and also indicate how those objectives were to be measured. This mandate (Rule C-1), known as CAPPS, is now the Educational Improvement Program. All school systems within the state complied with the mandate and, in most cases, utilized the skills and expertise of various teacher and lay advisory groups. The overall task was quite time-consuming at the local levels but tended to result in a far greater awareness by all concerned of what the major educational objectives within each school corporation are and should be.

More than 90 percent of the plans for curricular assessment reaching IDPI headquarters indicated measurement of attainment of local objectives would be done through the use of standardized achievement tests. Research conducted and reported by the Evaluation Assistance Center staff suggests the testing component for meeting the CAPPS mandate was not as carefully considered by public school personnel as was the component of developing and refining objectives. The EAC research, carried out in a number of school systems throughout Indiana, examined local curricular objectives and the objectives being tested for by the standardized achievement tests utilized in those school systems. In every case, less than a 50 percent overlap was found between stated local objectives and the objectives actually tested for with standardized tests. This in no way suggests that the local objectives were inappropriate nor should it be assumed that the standardized instruments were of low quality. Rather, the conclusion to be drawn is that it is most difficult to measure locally developed objectives with testing instruments that were designed to serve educators on a national basis.

To further "muddy the waters," the EAC 1982 opinionnaire results found that public school teachers and administrators, on the whole, do not feel they have adequate skills in using standardized achievement tests to measure pupil growth or to effect curricular change. So, the decision to go with standardized instruments to measure the locally developed CAPPS objectives appears to have been made without examining the ramifications of such a decision.

Teachers and administrators throughout Indiana indicated they felt that teacher skills in developing classroom tests far exceed their skills in the use of standardized achievement test results. Therefore, the objectives that were established at the local school corporation level could very well be measured by tests that are developed at the local corporation level. The initial workload on teachers and principals would be quite heavy in that it takes a great deal of time to build a good test, but the results would almost certainly provide information more usable than that obtained through standardized testing.

At the time of this writing, the Educational Improvement Program sponsored by the Indiana Department of Public Instruction appears to be in a "holding pattern." The original plan was to go through a second cycle, repeating (and refining) original efforts made at the local level. If school personnel are to carry out such a process, the staff of the Evaluation Assistance Center strongly suggests that the major emphasis be placed on the development of local tests to measure local objectives. A few schools in Indiana are doing this and are finding great success in further promoting school achievement. Public school administrators, as busy as they are, need to take the lead to promote quality measurement and evaluation in their respective schools.

Measurement and evaluation in the school setting should be an ever-evolving practice. It is as much of the teaching process as is the development of objectives and the teaching to those objectives. Without a quality measurement and evaluation program, it is most difficult to determine with any assuredness just how effective the teaching process has been. The Evaluation Assistance Center study conducted during the Spring of 1982 has provided data that suggest needs exist in the development and furthering of skills in measurement and evaluation throughout the state, whether those measures are made through classroom testing or through standardized testing. It is most encouraging to find the data also suggesting that teachers and administrators are aware of these needs and that they recognize them to be worthy of pursuit.

APPENDIX

TEACHER OPINIONNAIRE

THIS OPINIONNAIRE HAS BEEN DEVELOPED TO AID THE EVALUATION ASSISTANCE CENTER IN DETERMINING THE CURRENT STATUS OF TESTING, MEASUREMENT, AND EVALUATION IN INDIANA.

1. What grades or grades do you now teach?_____
2. What are the grades taught in your school? _____
3. List the primary subject or special area which your presently teach. (If self-contained, write "all.")_____
4. Have you taken a graduate class in educational measurement?_____

Directions:
For this column circle the number which corresponds to the word best describing the degree of **familiarity you feel you have** with each area under consideration.

Directions:
For this column circle the number which corresponds to the word best describing the degree of **familiarity you and your colleagues should have** with each area under consideration.

PLEASE RESPOND TO THIS OPINIONNAIRE. YOUR FEELINGS AND OPINIONS ARE IMPORTANT. AFTER RESPONDING, PLACE IN THE POSTAGE PAID ENVELOPE PROVIDED AND PLEASE RETURN.

	Familiarity you have					Familiarity you should have				
	thorough	adequate	minimal	none		thorough	adequate	minimal	none	
	4	3	2	1		4	3	2	1	
					REASONS FOR TESTING					
1	4	3	2	1	Motivating students to learn	4	3	2	1	2
3	4	3	2	1	Evaluation of student learning	4	3	2	1	4
5	4	3	2	1	Curriculum evaluation	4	3	2	1	6
7	4	3	2	1	Diagnosis of pupil strengths	4	3	2	1	8
					STANDARDIZED TESTING					
9	4	3	2	1	Justifying a standardized testing program	4	3	2	1	10
11	4	3	2	1	Using results in individual classrooms	4	3	2	1	12
13	4	3	2	1	Interpretation of results for school wide curriculum change	4	3	2	1	14
15	4	3	2	1	Sharing results with students	4	3	2	1	16
17	4	3	2	1	Sharing results with parents	4	3	2	1	18
19	4	3	2	1	Sharing results with the general public	4	3	2	1	20
21	4	3	2	1	Interpretation of specific scores (percentiles, grade equivalents, and stanines)	4	3	2	1	22
23	4	3	2	1	Understanding how much confidence can be placed in standardized test results	4	3	2	1	24
25	4	3	2	1	Selecting the proper test	4	3	2	1	26
27	4	3	2	1	Teaching test taking skills	4	3	2	1	28

FOR COMPUTER USE ONLY

FOR COMPUTER USE ONLY

FOR COMPUTER USE ONLY	Familiarity you have: thorough (4)	adequate (3)	minimal (2)	none (1)		Familiarity you should have: thorough (4)	adequate (3)	minimal (2)	none (1)	FOR COMPUTER USE ONLY
					TEACHER MADE TESTING					
29	4	3	2	1	Sharing test results with parents	4	3	2	1	30
31	4	3	2	1	Making the test blueprint	4	3	2	1	32
33	4	3	2	1	True-false tests	4	3	2	1	34
35	4	3	2	1	Multiple-choice tests	4	3	2	1	36
37	4	3	2	1	Matching tests	4	3	2	1	38
39	4	3	2	1	Classification tests	4	3	2	1	40
41	4	3	2	1	Short-answer tests	4	3	2	1	42
43	4	3	2	1	Completion tests	4	3	2	1	44
45	4	3	2	1	Essay tests	4	3	2	1	46
47	4	3	2	1	Analyzing teacher made tests	4	3	2	1	48
					TEST ENVIRONMENT					
49	4	3	2	1	Preparing students for test taking	4	3	2	1	50
51	4	3	2	1	Physical conditions appropriate for testing	4	3	2	1	52
53	4	3	2	1	Psychological factors affecting test results	4	3	2	1	54
					THE COMPUTER IN TESTING (micro-computer included)					
55	4	3	2	1	Using the computer to test pupils	4	3	2	1	56
57	4	3	2	1	Storing test data on computer disc or tape	4	3	2	1	58
59	4	3	2	1	Using the computer for test interpretation	4	3	2	1	60
61	4	3	2	1	Using the computer for data analysis	4	3	2	1	62
63	4	3	2	1	Retrieving pupil assessment data	4	3	2	1	64
					MISCELLANEOUS CONCERNS					
65	4	3	2	1	When mastery testing is appropriate	4	3	2	1	66
67	4	3	2	1	When it is best to "grade on the curve"	4	3	2	1	68
69	4	3	2	1	Measuring through checklists, grading scales, and scorecards	4	3	2	1	70
71	4	3	2	1	Improving learning and instruction through quality testing	4	3	2	1	72
73	4	3	2	1	Different methods of determining grades	4	3	2	1	74
75	4	3	2	1	Retention and promotion alternatives	4	3	2	1	76
77	4	3	2	1	Using tests to measure instructional effectiveness	4	3	2	1	78
79	4	3	2	1	Preparing students to take the SAT	4	3	2	1	80
81	4	3	2	1	Alternatives to the SAT	4	3	2	1	82

THANKS FOR RESPONDING.

PLEASE PLACE IN THE POSTAGE PAID ENVELOPE PROVIDED AND RETURN.

PRINCIPAL OPINIONNAIRE

THIS OPINIONNAIRE HAS BEEN DEVELOPED TO AID THE EVALUATION ASSISTANCE CENTER IN DETERMINING THE CURRENT STATUS OF TESTING, MEASUREMENT, AND EVALUATION IN INDIANA.

1. For how many years have you been a principal? _____
2. What are the grades taught in your school? _____
3. What is the approximate enrollment of your school? _____
4. Have you taken a graduate class in educational measurement? _____

	Directions: For this column circle the number which corresponds to the word best describing the degree of **familiarity you feel teachers should have** with each area under consideration.					Directions: For this column circle the number which corresponds to the word best describing the degree of **familiarity you should have** with each area under consideration.				
	Familiarity teachers should have				PLEASE RESPOND TO THIS OPINIONNAIRE. YOUR FEELINGS AND OPINIONS ARE IMPORTANT. AFTER RESPONDING, PLACE IN THE POSTAGE PAID ENVELOPE PROVIDED AND PLEASE RETURN.	Familiarity you should have				
	thorough	adequate	minimal	none		thorough	adequate	minimal	none	
	4	3	2	1		4	3	2	1	
					REASONS FOR TESTING					
1	4	3	2	1	Motivating students to learn	4	3	2	1	2
3	4	3	2	1	Evaluation of student learning	4	3	2	1	4
5	4	3	2	1	Curriculum evaluation	4	3	2	1	6
7	4	3	2	1	Diagnosis of pupil strengths	4	3	2	1	8
					STANDARDIZED TESTING					
9	4	3	2	1	Justifying a standardized testing program	4	3	2	1	10
11	4	3	2	1	Using results in individual classrooms	4	3	2	1	12
13	4	3	2	1	Interpretation of results for school wide curriculum change	4	3	2	1	14
15	4	3	2	1	Sharing results with students	4	3	2	1	16
17	4	3	2	1	Sharing results with parents	4	3	2	1	18
19	4	3	2	1	Sharing results with the general public	4	3	2	1	20
21	4	3	2	1	Interpretation of specific scores (percentiles, grade equivalents, and stanines)	4	3	2	1	22
23	4	3	2	1	Understanding how much confidence can be placed in standardized test results	4	3	2	1	24
25	4	3	2	1	Selecting the proper test	4	3	2	1	26
27	4	3	2	1	Teaching test taking skills	4	3	2	1	28

FOR COMPUTER USE ONLY

FOR COMPUTER USE ONLY

FOR COMPUTER USE ONLY	Familiarity teachers should have: thorough (4)	adequate (3)	minimal (2)	none (1)		Familiarity you should have: thorough (4)	adequate (3)	minimal (2)	none (1)	FOR COMPUTER USE ONLY
					TEACHER MADE TESTING					
29	4	3	2	1	Sharing test results with parents	4	3	2	1	30
31	4	3	2	1	Making the test blueprint	4	3	2	1	32
33	4	3	2	1	True-false tests	4	3	2	1	34
35	4	3	2	1	Multiple-choice tests	4	3	2	1	36
37	4	3	2	1	Matching tests	4	3	2	1	38
39	4	3	2	1	Classification tests	4	3	2	1	40
41	4	3	2	1	Short-answer tests	4	3	2	1	42
43	4	3	2	1	Completion tests	4	3	2	1	44
45	4	3	2	1	Essay tests	4	3	2	1	46
47	4	3	2	1	Analyzing teacher made tests	4	3	2	1	48
					TEST ENVIRONMENT					
49	4	3	2	1	Preparing students for test taking	4	3	2	1	50
51	4	3	2	1	Physical conditions appropriate for testing	4	3	2	1	52
53	4	3	2	1	Psychological factors affecting test results	4	3	2	1	54
					THE COMPUTER IN TESTING (micro-computer included)					
55	4	3	2	1	Using the computer to test pupils	4	3	2	1	56
57	4	3	2	1	Storing test data on computer disc or tape	4	3	2	1	58
59	4	3	2	1	Using the computer for test interpretation	4	3	2	1	60
61	4	3	2	1	Using the computer for data analysis	4	3	2	1	62
63	4	3	2	1	Retrieving pupil assessment data	4	3	2	1	64
					MISCELLANEOUS CONCERNS					
65	4	3	2	1	When mastery testing is appropriate	4	3	2	1	66
67	4	3	2	1	When it is best to "grade on the curve"	4	3	2	1	68
69	4	3	2	1	Measuring through checklists, grading scales, and scorecards	4	3	2	1	70
71	4	3	2	1	Improving learning and instruction through quality testing	4	3	2	1	72
73	4	3	2	1	Different methods of determining grades	4	3	2	1	74
75	4	3	2	1	Retention and promotion alternatives	4	3	2	1	76
77	4	3	2	1	Using tests to measure instructional effectiveness	4	3	2	1	78
79	4	3	2	1	Preparing students to take the SAT	4	3	2	1	80
81	4	3	2	1	Alternatives to the SAT	4	3	2	1	82

THANKS FOR RESPONDING.

PLEASE PLACE IN THE POSTAGE PAID ENVELOPE PROVIDED AND RETURN.

36

CENTRAL STAFF OPINIONNAIRE

1. What is your official title? _____

2. Curriculum responsibilities for what levels? (K-12, 9-12, etc.)? _____

3. Number of certified teaching staff in your corporation? _____

4. Have you taken a graduate class in educational measurement? _____

Directions:
For this column circle the number which corresponds to the word best describing the degree of **familiarity teachers** and/or **principals have** with each area under consideration.

Directions:
For this column circle the number which corresponds to the word best describing the degree of **familiarity teachers** and/or **principals should have** with each area under consideration.

PLEASE RESPOND TO THIS OPINIONNAIRE. YOUR FEELINGS AND OPINIONS ARE IMPORTANT. AFTER RESPONDING, PLACE IN THE POSTAGE PAID ENVELOPE PROVIDED AND PLEASE RETURN.

	Familiarity teachers/principals have					Familiarity teachers/principals should have				
	thorough	adequate	minimal	none		thorough	adequate	minimal	none	
	4	3	2	1		4	3	2	1	
					REASONS FOR TESTING					
1	4	3	2	1	Motivating students to learn	4	3	2	1	2
3	4	3	2	1	Evaluation of student learning	4	3	2	1	4
5	4	3	2	1	Curriculum evaluation	4	3	2	1	6
7	4	3	2	1	Diagnosis of pupil strengths	4	3	2	1	8
					STANDARDIZED TESTING					
9	4	3	2	1	Justifying a standardized testing program	4	3	2	1	10
11	4	3	2	1	Using results in individual classrooms	4	3	2	1	12
13	4	3	2	1	Interpretation of results for school wide curriculum change	4	3	2	1	14
15	4	3	2	1	Sharing results with students	4	3	2	1	16
17	4	3	2	1	Sharing results with parents	4	3	2	1	18
19	4	3	2	1	Sharing results with the general public	4	3	2	1	20
21	4	3	2	1	Interpretation of specific scores (percentiles, grade equivalents, and stanines)	4	3	2	1	22
23	4	3	2	1	Understanding how much confidence can be placed in standardized test results	4	3	2	1	24
25	4	3	2	1	Selecting the proper test	4	3	2	1	26
27	4	3	2	1	Teaching test taking skills	4	3	2	1	28

FOR COMPUTER USE ONLY

FOR COMPUTER USE ONLY

					TEACHER MADE TESTING					
29	4	3	2	1	Sharing test results with parents	4	3	2	1	30
31	4	3	2	1	Making the test blueprint	4	3	2	1	32
33	4	3	2	1	True-false tests	4	3	2	1	34
35	4	3	2	1	Multiple-choice tests	4	3	2	1	36
37	4	3	2	1	Matching tests	4	3	2	1	38
39	4	3	2	1	Classification tests	4	3	2	1	40
41	4	3	2	1	Short-answer tests	4	3	2	1	42
43	4	3	2	1	Completion tests	4	3	2	1	44
45	4	3	2	1	Essay tests	4	3	2	1	46
47	4	3	2	1	Analyzing teacher made tests	4	3	2	1	48
					TEST ENVIRONMENT					
49	4	3	2	1	Preparing students for test taking	4	3	2	1	50
51	4	3	2	1	Physical conditions appropriate for testing	4	3	2	1	52
53	4	3	2	1	Psychological factors affecting test results	4	3	2	1	54
					THE COMPUTER IN TESTING (micro-computer included)					
55	4	3	2	1	Using the computer to test pupils	4	3	2	1	56
57	4	3	2	1	Storing test data on computer disc or tape	4	3	2	1	58
59	4	3	2	1	Using the computer for test interpretation	4	3	2	1	60
61	4	3	2	1	Using the computer for data analysis	4	3	2	1	62
63	4	3	2	1	Retrieving pupil assessment data	4	3	2	1	64
					MISCELLANEOUS CONCERNS					
65	4	3	2	1	When mastery testing is appropriate	4	3	2	1	66
67	4	3	2	1	When it is best to "grade on the curve"	4	3	2	1	68
69	4	3	2	1	Measuring through checklists, grading scales, and scorecards	4	3	2	1	70
71	4	3	2	1	Improving learning and instruction through quality testing	4	3	2	1	72
73	4	3	2	1	Different methods of determining grades	4	3	2	1	74
75	4	3	2	1	Retention and promotion alternatives	4	3	2	1	76
77	4	3	2	1	Using tests to measure instructional effectiveness	4	3	2	1	78
79	4	3	2	1	Preparing students to take the SAT	4	3	2	1	80
81	4	3	2	1	Alternatives to the SAT	4	3	2	1	82

FOR COMPUTER USE ONLY

FOR COMPUTER USE ONLY

38

	Familiarity you have				Directions: For this column circle the number which corresponds to the word best describing the degree of familiarity you feel you have with each area under consideration.	Directions: For this column circle the number which corresponds to the word best describing the degree of familiarity you feel you should have with each area under consideration.	Familiarity you should have			
	thorough	adequate	minimal	none			thorough	adequate	minimal	none
	4	3	2	1			4	3	2	1
					USING STANDARDIZED TESTS FOR...					
83	4	3	2	1	Diagnosing curricular strengths		4	3	2	1
85	4	3	2	1	Identifying curricular weaknesses		4	3	2	1
87	4	3	2	1	Placement (tracking) of students		4	3	2	1
89	4	3	2	1	Determining teacher effectiveness		4	3	2	1
91	4	3	2	1	Determining principal effectiveness		4	3	2	1
93	4	3	2	1	Sharing program effectiveness with the general public		4	3	2	1
95	4	3	2	1	Justifying curricular change		4	3	2	1
					USING THE COMPUTER (MICROCOMPUTER INCLUDED) FOR...					
97	4	3	2	1	Making curricular projections		4	3	2	1
99	4	3	2	1	Scoring tests		4	3	2	1
101	4	3	2	1	Analyzing test results		4	3	2	1
103	4	3	2	1	Data storage and retrieval		4	3	2	1
105	4	3	2	1	Test score interpretation		4	3	2	1

FOR COMPUTER USE ONLY

THANKS FOR RESPONDING.

PLEASE PLACE IN THE POSTAGE PAID ENVELOPE PROVIDED AND RETURN.

39

Additional copies of this bulletin may be ordered from:

CURRICULUM RESEARCH AND DEVELOPMENT CENTER
Statesman Towers West
School of Education
Indiana State University
Terre Haute, Indiana 47809

Other bulletins currently available:

A PAGEANT IN READING
READING CHALLENGES: NEW DIRECTIONS
THE READING ENVIRONMENT
RHYTHM, RHYME, AND REASON
READING: FOCUS FOR ENJOYMENT
READING: GREAT EXPECTATIONS

CURRENT ISSUES IN SOCIAL STUDIES EDUCATION
RE-EDUCATION IN HUMAN COMMUNICATIONS
ACTIVITIES FOR ELEMENTARY SCHOOL MATHEMATICS ENRICHMENT
ALTERNATIVES TO TESTS, MARKS, AND CLASS RANK
LEARNING ACTIVITY PACKAGES

CURRICULAR TRENDS IN INDIANA PUBLIC SCHOOLS
ENVIRONMENTAL EDUCATION IN INDIANA PUBLIC SCHOOLS
PERCEPTIONS OF FUTURE ELEMENTARY SCHOOL PRINCIPALS
THE ROLE OF CLASSROOM TEACHERS IN STUDENT SELF-CONCEPT DEVELOPMENT
THE ARENAS OF POWER: FOCUS ON SCHOOLING

CURRICULUM EVOLUTION AS PORTRAYED THROUGH OLD TEXTBOOKS
MINORITY ETHNIC EDUCATION IN INDIANA PUBLIC SCHOOLS
BLACK PEOPLE IN AMERICA: A TEACHING-LEARNING UNIT
THE TEACHER'S ROLE IN OTHERS-CONCEPT DEVELOPMENT
THE DIFFERENCE BETWEEN THORNS AND THREATS IN ELEMENTARY SCHOOL ADMINISTRATION

THE THORNS AND THREATS OF SECONDARY SCHOOL ADMINISTRATION
A DEVELOPING CURRICULUM IN ENGLISH
INDIANA HISTORY TEST: FOURTH GRADE
THE AMERICAN INDIAN: A TEACHING-LEARNING UNIT
LEADERSHIP FOR EFFECTIVE EVALUATION OF LEARNING

CURRICULUM DEVELOPMENT IN INDIANA: ANALYSIS OF THE C-1 RULE AND CAPPS
TWELVE-YEAR FOLLOW-UP STUDY: ROLE AND STATUS OF CURRICULUM WORKERS IN INDIANA
USING MEASUREMENT & EVALUATION TO PROMOTE LEARNING
SOCIALLY MAINSTREAMING HANDICAPPED STUDENTS